

Research Article

Role of Attention and Perceptual Grouping in Visual Statistical Learning

Chris I. Baker,¹ Carl R. Olson,^{1,2,3} and Marlene Behrmann^{1,2,3}

¹Center for the Neural Basis of Cognition, Mellon Institute; ²Department of Neuroscience, University of Pittsburgh; and

³Department of Psychology, Carnegie Mellon University

ABSTRACT—Statistical learning has been widely proposed as a mechanism by which observers learn to decompose complex sensory scenes. To determine how robust statistical learning is, we investigated the impact of attention and perceptual grouping on statistical learning of visual shapes. Observers were presented with stimuli containing two shapes that were either connected by a bar or unconnected. When observers were required to attend to both locations at which shapes were presented, the degree of statistical learning was unaffected by whether the shapes were connected or not. However, when observers were required to attend to just one of the shapes' locations, statistical learning was observed only when the shapes were connected. These results demonstrate that visual statistical learning is not just a passive process. It can be modulated by both attention and connectedness, and in natural scenes these factors may constrain the role of stimulus statistics in learning.

It is well established that human observers learn auditory or visual patterns defined statistically or probabilistically, such as stimuli that co-occur frequently (Chun, 2002; Fiser & Aslin, 2001, 2002b; Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin, & Newport, 1999). Such learning, often called “statistical,” is commonly described as incidental or implicit, in that learning occurs automatically without instruction and without observers attending to the patterns explicitly. For example, statistical learning has been demonstrated both when stimuli are presented passively without any explicit task (Fiser & Aslin, 2001) and when observers are attending to and performing a separate, unrelated task (Saffran, Newport, Aslin, Tunick, & Barrueco, 1997). Such results have led to the view of statistical learning as a passive absorption of statistical regularities. Furthermore, statistical learning has been observed in both nonhuman primates (Hauser,

Newport, & Aslin, 2001) and human infants (Fiser & Aslin, 2002a; Kirkham, Slemmer, & Johnson, 2002; Saffran et al., 1996).

Statistical learning of visual shape combinations (Chun & Jiang, 1999; Edelman, Hiles, Yang, & Intrator, 2002; Fiser & Aslin, 2001) may provide a mechanism for generating object or scene representations. For example, Fiser and Aslin (2001) presented observers with a series of displays consisting of six shapes embedded in a three-by-three grid. Across displays, some shapes co-occurred frequently, whereas others did not. Following this exposure, participants could discriminate between frequent and infrequent shape pairs, a result suggesting that they had formed explicit representations of the stimulus statistics. Chun and Jiang (1999) also found frequency effects, with participants responding faster for frequent than infrequent shape pairings in a visual search task. However, because participants could not discriminate between frequent and infrequent pairings, Chun and Jiang suggested that representations formed by visual statistical learning are implicit.

Although these studies demonstrated robust statistical learning of visual shapes, they did not probe the nature of conditions under which learning occurs. In particular, they did not investigate whether statistical learning is sensitive to factors that, in other contexts, affect whether different shapes are processed together. The factors on which we focus here are attention and perceptual grouping.

Previous studies of statistical learning of shapes have not systematically manipulated attention, and the extent to which attention is even needed for statistical learning is controversial. Fiser and Aslin (2001) suggested that statistical learning is automatic once “general attention” is applied to a scene, and in the auditory domain, Saffran et al. (1996) observed statistical learning in the absence of directed attention. However, studies of statistical learning in other contexts have demonstrated learning for attended items only. For example, in a study of statistical learning of spatial location in visual search, Jiang and Chun (2001) found that observers learned statistical relations only between attended distractors and target location. Statistical learning of shapes may be similarly constrained by selective attention.

The impact of grouping cues on visual statistical learning has also not been examined. Increased perceptual binding of shapes might

Address correspondence to Chris I. Baker, McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, MIT, 77 Massachusetts Ave., NE20-443, Cambridge, MA 02139; e-mail: cibaker@mit.edu.

make them more susceptible to statistical learning. Because grouping cues are prevalent in real-world scenes, understanding their impact will cast light on the relevance of statistical learning for perception under natural circumstances.

In this study, we employed a paradigm allowing systematic control of both attention and grouping. The stimuli contained two shapes in fixed locations, and the frequency of co-occurrence of shapes was manipulated.

To investigate the role of attention, we asked observers to perform a task requiring attention to either one or both shape locations. Each stimulus contained one target shape and one distractor. Targets instructed either a left- or a right-lever response, and the observer's task was to find the target and make the appropriate response. In some experiments, targets could occupy either location (so that attention had to be directed to both), whereas in other experiments, targets occupied one location only (so that attention could be focused at this location). It is important to note that this task was orthogonal to the manipulation of stimulus statistics, and therefore any learning of the shape pairs was incidental.

To investigate the impact of grouping, we manipulated the connectedness of the shapes (either connected by a bar or separate). Connectedness is one of the strongest cues for visual grouping (Palmer & Rock, 1994) and facilitates the integration of visual parts in both shape perception (Saiki & Hummel, 1998b) and category learning (Saiki & Hummel, 1998a).

We used two measures of statistical learning. First, we measured task performance. Statistical learning would result in faster or more accurate performance for frequent than for infrequent shape pairs. Second, after subjects completed the task, we asked them to rate the familiarity of shape pairs. Statistical learning would result in higher familiarity ratings for frequent than infrequent pairs. Use of these different measures allowed us to determine whether the representations of stimulus statistics were implicit or explicit. Implicit representations would be reflected in differential task performance for frequent and infrequent shape pairs in the absence of differential familiarity judgments. Differences in ratings for frequent and infrequent pairs, however, would suggest that the representations were explicit.

The results indicate that when participants attended to both shape locations, statistical learning occurred regardless of connectedness. However, when participants attended to one location only, statistical learning occurred only when the shapes were connected.

GENERAL METHOD

All experiments used the general procedure outlined in this section. Deviations from this procedure are described for each experiment.

Design

Each stimulus contained two shapes (Fig. 1), which were either connected by a bar (Experiments 1 and 3) or separate (Experiments 2 and 4). Shapes were approximately 0.75° of visual angle high, and the total height of the stimuli was 2.5° .

In each experiment, there were eight target shapes and eight distractor shapes. Stimuli were constructed by combining one target and one distractor, for a total of 32 stimuli. Each target was associated with a given response (left or right). Distractors were paired equally often

with targets eliciting left and right responses and carried no information about response.

The critical variable was frequency of co-occurrence of target-distractor combinations. Frequent combinations were presented four times as often as infrequent combinations. Stimuli were divided into two sets. For a given participant, one set was designated "frequent" and the other "infrequent," with designations counterbalanced across participants.

Stimulus presentation and data collection were controlled using Cortex software (National Institute of Mental Health, Laboratory of Neuropsychology, <http://www.cortex.salk.edu/>).

Task

At the start of each trial, a fixation point was presented at the center of the screen, and participants depressed two levers. The fixation point turned red and after 500 ms was replaced by a stimulus for 100 ms. Four targets instructed a left response and four instructed a right response. Participants responded by releasing the appropriate lever. On each trial, feedback was given: three tones for a correct response and a red circle for an incorrect response. Participants were told the nature of the task, but were not informed of the response mappings of the targets. Practice trials on a separate stimulus set were given before the main experiment.

Each experiment consisted of 800 trials presented in 10 blocks, each lasting approximately 4 min. After 5 blocks, participants were given a short break before resuming. Within each block, frequent target-distractor combinations were presented four times, and infrequent combinations once.

After the experiment, participants were presented with 40 stimuli (all 32 experimental and 8 novel stimuli) and asked to rate them for familiarity on a 5-point scale (1 = *least familiar*, 5 = *most familiar*). In Experiments 1 and 2, novel stimuli were either target-target or distractor-distractor combinations, and subjects could identify these as novel if they noticed the presence of two targets or the absence of any. In Experiments 3 and 4, novel stimuli were new target-distractor combinations and thus were harder to identify as novel.

Analysis

The experiments were run in a three-factor design with frequency (high or low), block (1–10), and set (frequent stimuli = Set 1 or Set 2) as factors. In all experiments, there were no significant effects involving set, and so the data are collapsed across this factor. Performance data were analyzed using repeated measures analysis of variance (ANOVA) with Greenhouse-Geisser correction where appropriate. Correct trials only were included for reaction time (RT) analysis. Ratings were analyzed using matched-pairs *t* tests with Bonferroni correction.

EXPERIMENT 1

The aim of Experiment 1 was to determine whether statistical learning would occur under conditions requiring participants to attend to both shape locations and, if so, to ascertain whether the underlying representations were implicit or explicit. The shapes were connected, and targets could occupy either location (Fig. 1a). Participants were not informed which shapes were targets and which distractors.

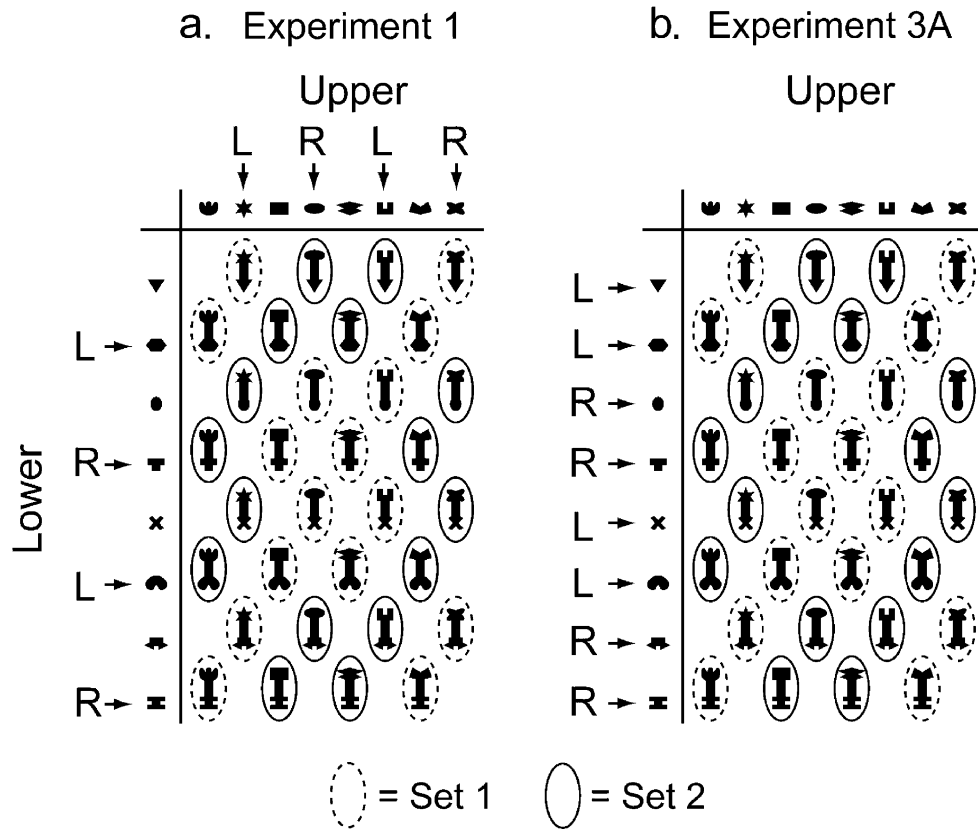


Fig. 1. Design and stimuli. In Experiment 1 (a), the stimulus set was constructed from eight upper shapes and eight lower shapes, shown at the top and left side of the grid. Each stimulus consisted of one upper shape connected by a bar to one lower shape, as shown within the grid. Four upper shapes and four lower shapes were designated as targets. The eight targets are indicated here by arrows and labeled with the designated response (R = right, L = left). The remaining eight shapes were distractors and were equally associated with left and right responses. The stimuli were equally divided into two sets (distinguished here by solid and dotted ovals); one was assigned to be presented with high frequency and the other with low frequency. In Experiment 2, the target and distractor shapes and response contingencies were identical, but there was no bar connecting the upper and lower shapes. In Experiment 3a (b), all targets occupied the lower location and all distractors, the upper location. As in Experiments 1 and 2, the distractors were equally associated with left and right responses. In Experiment 3b, the targets and distractors were reversed. In Experiment 4, targets and distractors were the same as in Experiment 3a, but there was no bar connecting the upper and lower shapes.

Participants

Twenty-four undergraduate and graduate students from Carnegie Mellon University (CMU) and the University of Pittsburgh participated for course credit, payment, or both. All participants had normal or corrected-to-normal vision. Familiarity ratings were obtained from 22 of the 24 participants.

Results

Over the course of the experiment, participants showed decreasing RTs and increasing accuracy for both frequent and infrequent target-distractor pairs. Critically, participants were faster and more accurate for frequent than for infrequent pairs (Fig. 2a). A Frequency \times Block repeated measures ANOVA revealed a main effect of frequency on RT, $F(1, 23) = 16.00$, $p < .0006$, $\eta_p^2 = .41$, and a marginal effect on accuracy, $F(1, 23) = 4.09$, $p < .055$, $\eta_p^2 = .15$. The effect of block was significant for accuracy, $F(3.2, 73.6) = 72.51$, $p < .00001$, $\eta_p^2 = .76$, but not for RT, $F(2.63, 60.5) = 2.3$, $p > .09$, $\eta_p^2 = .09$. There were no interactions of block and frequency (both $ps > .1$). These results

demonstrate statistical learning of target-distractor combinations. The lack of a Frequency \times Block interaction suggests that learning may be rapid, although the increased variability in early blocks due to the small percentage of correct trials makes the speed of learning difficult to assess.

Participants rated frequent stimuli as more familiar than infrequent stimuli (Fig. 2a). A direct comparison of these ratings using a matched-pairs t test revealed a significant effect of frequency, $t(21) = 3.02$, $p < .021$, $d = 0.40$. Ratings for both frequent and infrequent stimuli were significantly greater than those for novel stimuli, both $ts(21) > 6.5$, $p < .001$, $d > 2.0$. These results confirm statistical learning of target-distractor combinations and indicate that the underlying representations were explicit.

EXPERIMENT 2

The aim of Experiment 2 was to investigate whether statistical learning would persist when cues favoring grouping of the shapes were

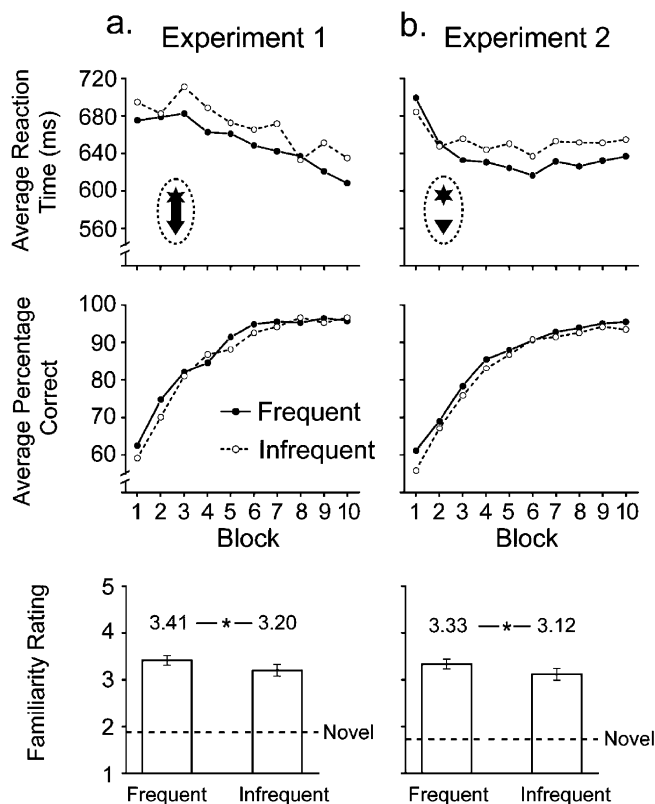


Fig. 2. Results from Experiment 1 (a) and Experiment 2 (b). Mean reaction time and accuracy are shown as a function of block, separately for the frequent and infrequent sets. Targets could appear at either the upper or the lower location, and participants were required to attend to both locations (as illustrated by the dotted ovals). The histograms at the bottom of the figure show mean familiarity ratings for the frequent and infrequent sets (the dashed lines indicate mean ratings for novel stimuli). Asterisks indicate significant differences between the ratings for frequent and infrequent items, $p < .025$.

attenuated. Procedures were identical to those of Experiment 1 with the exception that there was no bar connecting the shapes.

Participants

Twenty-four undergraduate and graduate students from CMU and the University of Pittsburgh participated for course credit, payment, or both. All participants had normal or corrected-to-normal vision.

Results

As in Experiment 1, participants were faster and more accurate for frequent than infrequent target-distractor pairs (Fig. 2b). A Frequency \times Block repeated measures ANOVA revealed a main effect of frequency on both RT, $F(1, 23) = 8.36$, $p < .009$, $\eta_p^2 = .27$, and accuracy, $F(1, 23) = 9.52$, $p < .006$, $\eta_p^2 = .29$. The effect of block was significant for accuracy, $F(3.3, 75.3) = 66.0$, $p < .00001$, $\eta_p^2 = .74$, but not for RT, $F(2.1, 48.0) = 0.96$, $p > .39$, $\eta_p^2 = .04$. There were no interactions of block and frequency (both $ps > .1$). These results indicate that even when the shapes were separated, participants showed statistical learning of the target-distractor combinations.

To compare these results with those of Experiment 1, we carried out a Connectedness \times Frequency \times Block repeated measures ANOVA.

This analysis revealed no effect of connectedness on either RT, $F(1, 46) = 0.34$, $p > .55$, $\eta_p^2 < .01$, or accuracy, $F(1, 46) = 1.41$, $p > .24$, $\eta_p^2 = .03$, and no interactions involving connectedness (all $ps > .3$). These results suggest that connectedness has no impact on statistical learning.

As in Experiment 1, participants rated frequent stimuli as more familiar than infrequent stimuli, $t(23) = 3.21$, $p < .012$, $d = 0.38$. Ratings for both frequent and infrequent stimuli were significantly greater than ratings for novel stimuli, both $ts(23) > 7.0$, $p < .001$, $d > 2.0$. A Connectedness (connected, unconnected) \times Frequency repeated measures ANOVA showed no effect of connectedness on ratings, $F(1, 44) = 0.27$, $p > .6$, $\eta_p^2 < .01$. These results confirm that there was no effect of connectedness on statistical learning and show that the representations were explicit.

EXPERIMENT 3

The aim of Experiment 3 was to investigate the role of attention in statistical learning. In Experiments 1 and 2, targets occupied the two possible shape locations with equal frequency. Allocation of attention to both locations might have favored the learning of shape combinations. In Experiment 3, all targets occupied one location only (Fig. 1b). This was the lower location in Experiment 3a, and the upper location in Experiment 3b. The stimuli were identical to those used in Experiment 1. Only the stimulus-response contingencies differed. Participants were told the location of targets and instructed to ignore the other location.

Participants

Thirty-six CMU undergraduate students (24 in Experiment 3a and 12 in Experiment 3b) participated for course credit, payment, or both. All participants had normal or corrected-to-normal vision.

Results

In Experiment 3a, when targets occupied the lower location, participants were faster responding to frequent than infrequent target-distractor combinations (Fig. 3a). A Frequency \times Block repeated measures ANOVA revealed a main effect of frequency on RT, $F(1, 23) = 21.22$, $p < .0002$, $\eta_p^2 = .48$, but not on accuracy, $F(1, 23) < 0.01$, $p > .95$, $\eta_p^2 < .01$. The main effect of block was significant for accuracy, $F(2.54, 58.32) = 35.11$, $p < .00001$, $\eta_p^2 = .60$, but not for RT, $F(3.3, 75.5) = 0.71$, $p > .55$, $\eta_p^2 = .03$, and there were no significant interactions (both $ps > .45$). These results indicate that even when participants were required to attend to one location only, statistical learning of target-distractor combinations persisted.

To compare the results with those of Experiment 1, we carried out a Task \times Frequency \times Block repeated measures ANOVA on both RT and accuracy. This revealed a significant main effect of task on RT, $F(1, 46) = 5.66$, $p < .022$, $\eta_p^2 = .11$, and a marginal effect on accuracy, $F(1, 46) = 3.94$, $p < .055$, $\eta_p^2 = .08$. The interaction of task and block was significant for accuracy, $F(3.1, 141.5) = 5.98$, $p < .0007$, $\eta_p^2 = .12$. These results indicate better performance overall and faster learning when participants were required to attend to one location only. The difference in RT between frequent and infrequent shape pairs was smaller in Experiment 3a than in Experiment 1, and there

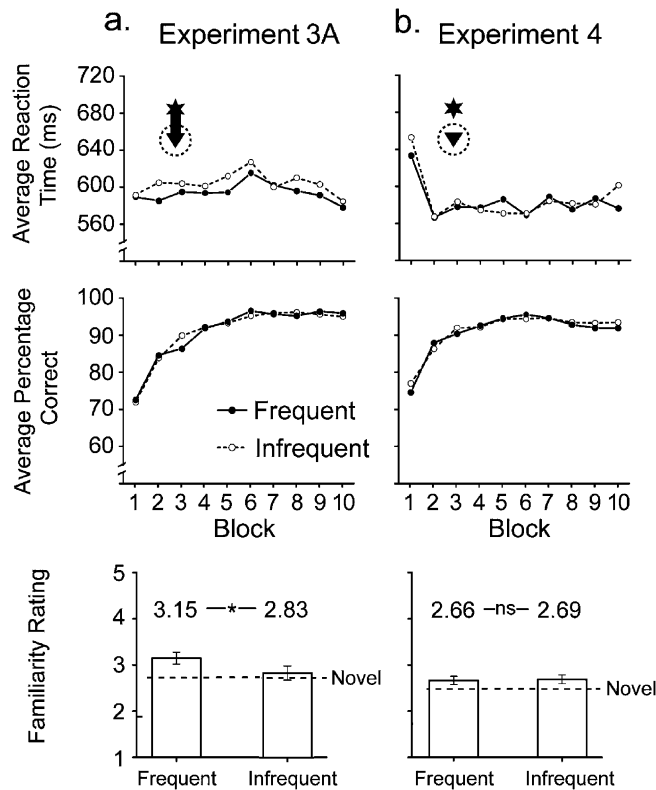


Fig. 3. Results from Experiment 3a (a) and Experiment 4 (b). Mean reaction time and accuracy are shown as a function of block, separately for the frequent and infrequent sets. Targets were always presented at the lower location, and participants were required to attend to this location only (as illustrated by the dotted circles). The histograms at the bottom of the figure show mean familiarity ratings for the frequent and infrequent sets (the dashed lines indicate mean ratings for novel stimuli). The asterisk indicates a significant difference between the ratings for frequent and infrequent items, $p < .01$.

was a trend for a Frequency \times Task interaction, $F(1, 46) = 3.13$, $p < .085$, $\eta_p^2 = .06$.

Participants rated frequent stimuli as more familiar than infrequent stimuli, $t(23) = 3.32$, $p < .009$, $d = 0.45$. Ratings for frequent stimuli were significantly greater than ratings for novel stimuli, $t(23) = 3.93$, $p < .003$, $d = 0.54$, but there was no difference between infrequent and novel stimuli, $t(23) = 1.17$, $p > .5$, $d = 0.11$. These results indicate that, as in the previous experiments, the representations of stimulus statistics were explicit.

In Experiment 3b, to ensure that there was no effect of the specific target location, we ran 12 new participants with targets occupying the upper location only. For this group of participants, one frequency assignment (i.e., designation of one set as high frequency and the other as low frequency) only was used. A Frequency \times Block repeated measures ANOVA revealed a main effect of frequency on RT, $F(1, 11) = 5.87$, $p < .034$, $\eta_p^2 = .35$, but not on accuracy, $F(1, 11) = 0.12$, $p > .73$, $\eta_p^2 = .10$. There was a significant main effect of block on accuracy, $F(2.7, 29.6) = 16.55$, $p < .00001$, $\eta_p^2 = .60$, and a marginal effect on RT, $F(1.8, 19.7) = 3.45$, $p < .057$, $\eta_p^2 = .24$. These results indicate statistical learning of target-distractor combinations.

To compare the results obtained in Experiments 3a and 3b, we performed a Target Location \times Frequency \times Block mixed ANOVA on

RT using participants run on the same frequency assignment (i.e., the one used for all Experiment 3b participants). This ANOVA revealed no effect of location, $F(1, 22) = 0.49$, $p > .49$, $\eta_p^2 = .02$, but a highly significant effect of frequency, $F(1, 22) = 16.10$, $p < .0006$, $\eta_p^2 = .42$. These results indicate that the specific locus of attention did not influence statistical learning.

EXPERIMENT 4

Observers in Experiment 3 might have processed both shapes, even though only one was relevant to the task, as a result of grouping processes called into play by the shapes' connectedness. The aim of Experiment 4 was to investigate this possibility. The procedure was identical to that of Experiment 3a with the exception that the bar connecting the shapes was removed.

Participants

Twenty-four CMU undergraduate students participated for course credit, payment, or both. All participants had normal or corrected-to-normal vision.

Results

Unlike in all the previous experiments, participants were neither faster nor more accurate for frequent than infrequent target-distractor combinations (Fig. 3b). A Frequency \times Block repeated measures ANOVA revealed no main effect of frequency on RT, $F(1, 23) = 0.90$, $p > .35$, $\eta_p^2 = .04$, or accuracy, $F(1, 23) = 0.84$, $p > .37$, $\eta_p^2 = .04$. There was, however, a significant main effect of block on both RT, $F(2.3, 53.4) = 3.11$, $p < .046$, $\eta_p^2 = .12$, and accuracy, $F(3.6, 82.9) = 24.2$, $p < .00001$, $\eta_p^2 = .51$, and a marginal interaction of frequency and block on RT, $F(5.0, 114.9) = 2.25$, $p < .055$, $\eta_p^2 = .09$. These results indicate that when participants were required to attend to one location only and the shapes at the two locations were not connected, there was no statistical learning of target-distractor combinations.

To compare these results with those obtained in Experiment 3a, we performed a Connectedness \times Frequency \times Block repeated measures ANOVA on both RT and accuracy. For accuracy, there were no significant effects involving connectedness (all $ps > .09$). For RT, there was a marginal interaction of block and connectedness, $F(3.1, 142.2) = 2.57$, $p < .055$, $\eta_p^2 = .05$, and a trend toward a Frequency \times Connectedness interaction, $F(1, 46) = 3.14$, $p < .085$, $\eta_p^2 = .06$.

Unlike in the earlier experiments, participants gave similar familiarity ratings to frequent and infrequent stimuli. There were no significant differences between the ratings for frequent, infrequent, and novel stimuli in matched-pairs t tests, all $ts(23) < 2.6$, $p > .05$. These results confirm that there was no statistical learning of the target-distractor combinations.

To compare these ratings with those in Experiment 3a, we carried out a Connectedness \times Frequency repeated measures ANOVA. This analysis revealed a highly significant Frequency \times Connectedness interaction, $F(1, 46) = 7.51$, $p < .009$, $\eta_p^2 = .14$. This result indicates a significant difference in statistical learning between Experiments 3a and 4.

GENERAL DISCUSSION

We have demonstrated that statistical learning is not a passive process, but is modulated by both attention and grouping in an interactive manner. When participants were required to attend to both locations at which shapes were presented, the degree of statistical learning was identical regardless of the connectedness of the shapes. This suggests that connectedness alone affords no extra benefit to statistical learning. When participants were required to attend to one location only, however, statistical learning was observed only when the shapes were connected. This suggests that, in the absence of explicit attention to two locations, connectedness affords greater sensitivity to statistical learning.

Statistical learning was evident in both measures of performance and familiarity ratings, suggesting that representations of shape combinations are explicit. In contrast, Chun and Jiang (1999) found evidence for implicit memory of statistical relations (RT in visual search) in the absence of explicit memory (familiarity judgments). In their study, however, the stimuli contained multiple distractors and a single target, and targets and distractors were not presented in fixed locations. It may be easier to form explicit representations of statistical relations among shapes when there are relatively few shapes presented in fixed positions. This would explain why Fiser and Aslin (2001), who also presented a small number of stimuli in fixed relative positions, also observed explicit memory for statistically defined patterns.

Previous studies have demonstrated statistical learning of unconnected visual shapes (Chun & Jiang, 1999; Edelman et al., 2002; Fiser & Aslin, 2001). Critically, we have shown that such learning will occur only if observers attend to all shape locations. Attention was not manipulated in earlier studies. For example, Fiser and Aslin (2001) gave no instructions to participants as to what to attend to in the stimuli, but it is likely that they attended to all shapes. Our finding is consistent with Jiang and Chun's (2001) finding for statistical learning of spatial location and suggests that statistical learning of visual shapes is constrained by attention in the absence of other grouping cues binding the shapes together.

The presence of statistical learning for connected but not unconnected shapes when participants were required to attend to one location suggests that connectedness promotes the binding of individual shapes in the absence of voluntary attention. The uniform connection between the stimuli may have promoted the automatic spreading of attention to the unattended location, a process that has been referred to as object-based attention (e.g., Behrmann, Zemel, & Mozer, 1998; Egly, Driver, & Rafal, 1994; Moore, Yantis, & Vaughan, 1998).

Previous studies of early perceptual learning have suggested that attention is necessary for learning to occur (Ahissar & Hochstein, 1993; Gilbert, Ito, Kapadia, & Westheimer, 2000). In such studies, learning is likely to involve early visual areas, such as V1 and V2, and extensive training and feedback is usually required. In contrast, it is likely that the learning in a task such as ours occurs at a higher level in the visual system (Fiser & Aslin, 2001). Furthermore, the learning we observed was incidental and occurred relatively rapidly (within a single session). Our results suggest that attention is also required for this high-level statistical learning. Because the learning we observed was incidental, our results further suggest that learning is automatic for attended items, and that attention alone is sufficient for the binding of visual shapes.

Our findings suggest that participants may have encoded the upper and lower shapes as a single object and formed unitary representations of the stimuli. Evidence for the formation of unitary representations of multiple features has also been obtained in explicit discrimination tasks requiring attention to multiple object features in order to perform above chance (Gauthier & Tarr, 1997; Goldstone, 2000; Shiffrin & Lightfoot, 1997). For example, Goldstone (2000) found pronounced improvements over time in a categorization task when the task required identifying the conjunction of five line segments, but not when the task could be solved by attending to just one line segment. He argued that the improvement occurred because of "unitization" of the individual line segments. Consonant with our results, such findings suggest that attention may be critical in building unitary representations of complex stimuli.

Recent monkey neurophysiological data suggest a neural mechanism by which such unitization might occur. Training monkeys on a discrimination task was found to increase the number of neurons in inferotemporal cortex coding the conjunction of visual features (Baker, Behrmann, & Olson, 2002).

An alternative interpretation of our findings is that participants were not building unitary representations, but were learning shape associations. Each distractor was paired frequently with two targets, and infrequently with two other targets. Thus, although distractors were not predictive of response, they were predictive of target identity, so that reduced RT for frequent pairings could reflect priming of target representations by the distractors. A possible neural mechanism is suggested by the finding that neurons in monkey inferotemporal cortex exhibit pair coding for visual stimuli learned in a paired-associate task (Messinger, Squire, Zola, & Albright, 2001; Sakai & Miyashita, 1991).

In conclusion, we have shown that visual statistical learning, although a robust learning mechanism, is not just a passive process, but is modulated by both attention and grouping. Our results are consistent with those from studies on language learning (Johnson & Jusczyk, 2001) in demonstrating that learners are sensitive to factors other than statistical regularities. The results suggest that attention to individual shapes is required for statistical learning, although this may be produced voluntarily by explicit direction of attention (top-down) or by the automatic spreading of attention induced by perceptual grouping (bottom-up). The deployment of attention thus constrains the extent to which statistical learning provides a mechanism for generating unitary object representations.

REFERENCES

- Ahissar, M., & Hochstein, S. (1993). Attentional control of early perceptual learning. *Proceedings of the National Academy of Sciences, USA*, 90, 5718–5722.
- Baker, C.I., Behrmann, M., & Olson, C.R. (2002). Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nature Neuroscience*, 5, 1210–1216.
- Behrmann, M., Zemel, R.S., & Mozer, M.C. (1998). Object-based attention and occlusion: Evidence from normal participants and a computational model. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1011–1036.
- Chun, M. (2002). Contextual cueing of visual attention. *Trends in Cognitive Sciences*, 4, 170–178.
- Chun, M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, 10, 360–365.

- Edelman, S., Hiles, B.P., Yang, H., & Intrator, N. (2002). Probabilistic principles in unsupervised learning of visual structure: Human data and a model. In T.G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Advances in neural information processing systems*, 14 (pp. 19–26). Cambridge, MA: MIT Press.
- Egely, R., Driver, J., & Rafal, R.D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, 123, 161–177.
- Fiser, J., & Aslin, R.N. (2001). Unsupervised learning of higher-order spatial structures from visual scenes. *Psychological Science*, 12, 499–504.
- Fiser, J., & Aslin, R.N. (2002a). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 458–467.
- Fiser, J., & Aslin, R.N. (2002b). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences, USA*, 99, 15822–15826.
- Gauthier, I., & Tarr, M.J. (1997). Becoming a “Greeble” expert: Exploring mechanisms for face recognition. *Vision Research*, 37, 1673–1682.
- Gilbert, C., Ito, M., Kapadia, M., & Westheimer, G. (2000). Interactions between attention, context and learning in primary visual cortex. *Vision Research*, 40, 1217–1226.
- Goldstone, R. (2000). Unitization during category learning. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 86–112.
- Hauser, M.D., Newport, E.L., & Aslin, R.N. (2001). Segmentation of the speech stream in a nonhuman primate: Statistical learning in cotton-top tamarins. *Cognition*, 78, B53–B64.
- Jiang, Y., & Chun, M.M. (2001). Selective attention modulates implicit learning. *Quarterly Journal of Experimental Psychology*, 54A, 1105–1124.
- Johnson, E.K., & Jusczyk, P.W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, 44, 548–567.
- Kirkham, N.Z., Slemmer, J.A., & Johnson, S.P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, 83, B35–B42.
- Messinger, A., Squire, L.R., Zola, S.M., & Albright, T.D. (2001). Neuronal representations of stimulus associations develop in the temporal lobe during learning. *Proceedings of the National Academy of Sciences, USA*, 98, 12239–12244.
- Moore, C.M., Yantis, S., & Vaughan, B. (1998). Object-based visual selection: Evidence from perceptual completion. *Psychological Science*, 9, 104–110.
- Palmer, S.E., & Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin & Review*, 1, 29–55.
- Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928.
- Saffran, J.R., Johnson, E.K., Aslin, R.N., & Newport, E.L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70, 27–52.
- Saffran, J.R., Newport, E.L., Aslin, R.N., Tunick, R.A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, 8, 101–105.
- Saiki, J., & Hummel, J.E. (1998a). Connectedness and part-relation integration in shape category learning. *Memory & Cognition*, 26, 1138–1156.
- Saiki, J., & Hummel, J.E. (1998b). Connectedness and the integration of parts with relations in shape perception. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 227–251.
- Sakai, K., & Miyashita, Y. (1991). Neural organization for the long-term memory of paired associates. *Nature*, 354, 152–155.
- Shiffrin, R.M., & Lightfoot, N. (1997). Perceptual learning of alphanumeric-like characters. In R.L. Goldstone, P.G. Schyns, & D.L. Medin (Eds.), *The psychology of learning and motivation: Vol. 36* (pp. 45–82). San Diego, CA: Academic Press.

(RECEIVED 4/9/03; REVISION ACCEPTED 6/20/03)